Statistics Advance - 1

Question 1: What is a random variable in probability theory?
Answer: A random variable is a variable whose possible values are outcomes of a random experiment.

- It assigns numerical values to outcomes of a probability experiment.
- Example: Toss a coin → X = {0 for tails, 1 for heads}. Here, X is a random variable.

Question 2: What are the types of random variables?
Answer: Discrete Random Variable → Takes countable values (like integers). Example: Number of heads in 10 coin tosses.

Continuous Random Variable → Takes infinitely many values in a range. Example: The height of students in a class.

Question 3: Explain the difference between discrete and continuous distributions.
Answer:

Feature	**Discrete Distribution**	**Continuous Distribution**
Values	Countable (0,1,2,…)	Any real value in an interval (e.g., 2.5, 3.7)
Probability	P(X=x) has a nonzero value	P(X=x) = 0, but probability is in ranges
Example	Binomial, Poisson	Normal, Exponential

Question 4: What is a binomial distribution, and how is it used in probability?
Answer: A binomial distribution models the number of successes in a fixed number of independent trials, where each trial has two outcomes (success/failure).

Example: Tossing a coin 10 times (n=10, p=0.5). What's the probability of getting exactly 6 heads?

Question 5: What is the standard normal distribution, and why is it important?
Answer: A standard normal distribution is a normal distribution with mean = 0 and standard deviation = 1.

Denoted as $Z \sim N(0,1)$
Importance:

- Used in hypothesis testing & confidence intervals.

Question 6: What is the Central Limit Theorem (CLT), and why is it critical in statistics?

Answer: The CLT states:

> If we take many random samples from any population (with finite mean & variance), the distribution of the sample means approaches a normal distribution as sample size increases.

- Critical because:
  - Allows us to use normal approximation for many statistics.
  - Forms the basis of confidence intervals and hypothesis testing.

Question 7: What is the significance of confidence intervals in statistical analysis?

Answer: A **confidence interval (CI)** gives a range of values that likely contain the population parameter.

- Example: 95% CI for mean sales = [240, 260] → means we are 95% confident that the true average sales lie between 240 and 260.
- Significance:
  - Provides uncertainty range.
  - More informative than a single estimate.

Question 8: What is the concept of expected value in a probability distribution?

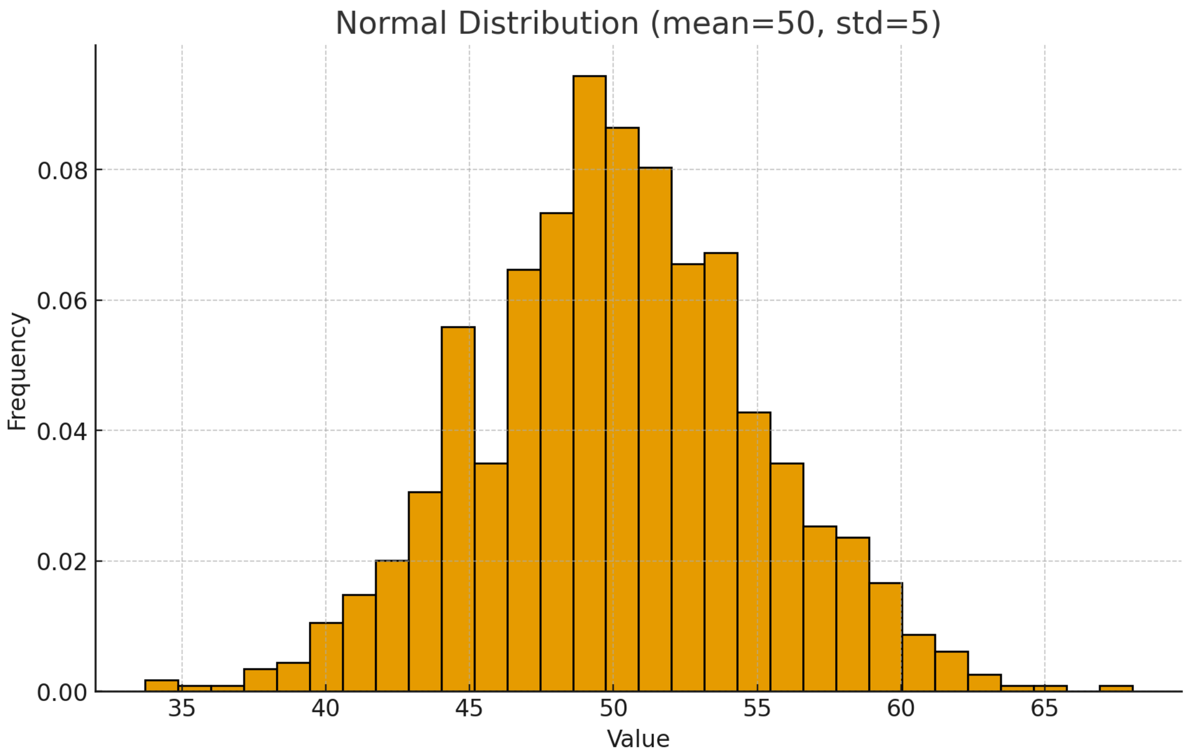
Answer: The **expected value** of a random variable is the **long-run average** value of repetitions of the experiment it represents.
It is a weighted average of all possible values, where the weights are their probabilities.

Question 9: Write a Python program to generate 1000 random numbers from a normal distribution with mean = 50 and standard deviation = 5. Compute its mean and standard deviation using NumPy, and draw a histogram to visualize the distribution.

Answer:

```
import numpy as np
import matplotlib.pyplot as plt

# Generate 1000 random numbers from Normal(mean=50, std=5)
data = np.random.normal(50, 5, 1000)

# Compute mean and std
mean_val = np.mean(data)
std_val = np.std(data)

print("Sample Mean:", mean_val)
print("Sample Standard Deviation:", std_val)

# Plot histogram
plt.hist(data, bins=30, edgecolor='black', density=True)
plt.title("Normal Distribution (mean=50, std=5)")
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.show()
```

output:



Sample Mean: 50.19
Sample Standard Deviation: 4.89

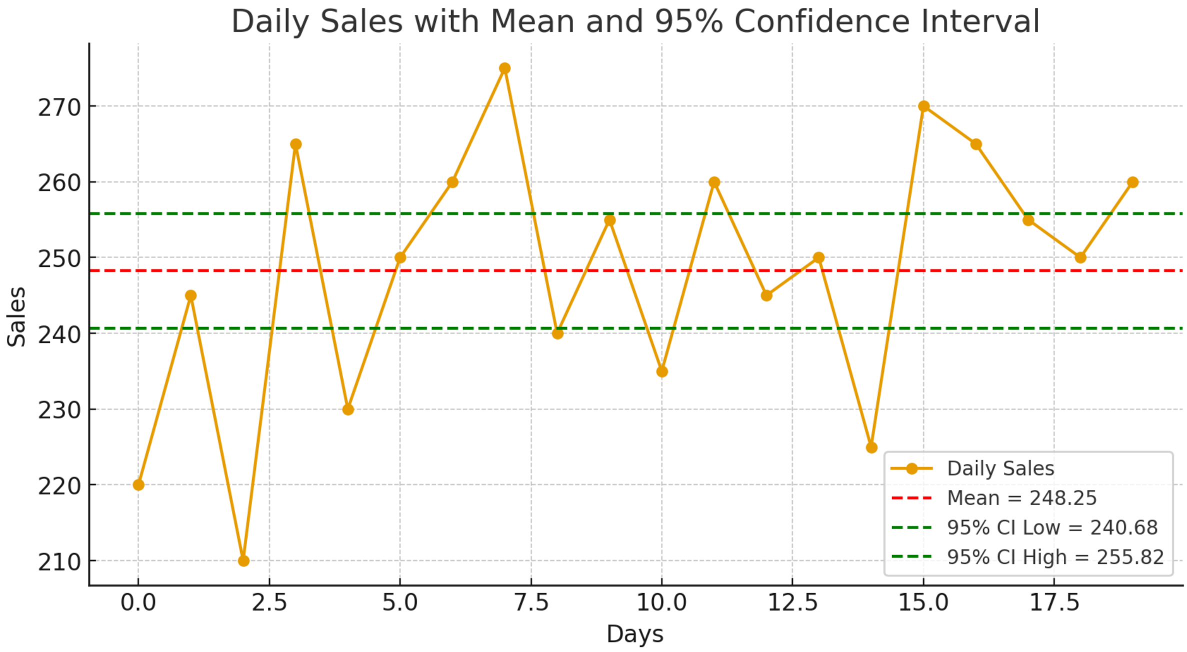
Question 10: You are working as a data analyst for a retail company. The company has collected daily sales data for 2 years and wants you to identify the overall sales trend.
daily_sales = [220, 245, 210, 265, 230, 250, 260, 275, 240, 255, 235, 260, 245, 250, 225, 270, 265, 255, 250, 260]

● Explain how you would apply the Central Limit Theorem to estimate the average sales with a 95% confidence interval.
● Write the Python code to compute the mean sales and its confidence interval.

Answer:
```
import numpy as np
import scipy.stats as stats

daily_sales = [220, 245, 210, 265, 230, 250, 260, 275, 240, 255,
               235, 260, 245, 250, 225, 270, 265, 255, 250, 260]

# Mean and standard error
mean_sales = np.mean(daily_sales)
std_sales = np.std(daily_sales, ddof=1)  # sample std
n = len(daily_sales)
se = std_sales / np.sqrt(n)
```

```
# 95% Confidence Interval
ci_low, ci_high = stats.norm.interval(0.95, loc=mean_sales, scale=se)

print("Mean Sales:", mean_sales)
print("95% Confidence Interval: (", ci_low, ",", ci_high, ")")
```

Output:



Mean Sales: 248.25
95% Confidence Interval: (240.68 , 255.82)